Comments from the Technical Advisory Committee on draft report: *Assessment of Nutrient Loading and Eutrophication in Barnegat Bay-Little Egg Harbor, New Jersey in Support of Nutrient Management Planning*

September 4, 2012

Overall Comments

General Comments

The USGS report is included as an Appendix to the report but its conclusions and recommendations are not clearly integrated or presented in the overall project final report.

An important expectation of this report was to determine the degree to which various stressors are responsible for the observed condition of Barnegat Bay. In addition, if possible, the threshold levels for each relevant stressor that would result in an impaired vs. unimpaired condition. To be useful in crafting a management strategy, it is essential to know both the relative importance of the suite of stressors that are responsible for the observed biotic condition, and the threshold for each stressor that aligns with an acceptable biotic condition. The elements of an effective management plan would depend on 1) which factors have the greatest influence on biotic condition, 2) which ones are most outside of the range needed for an acceptable biotic condition and 3) which factors can be remediated most cost effectively.

Overall, our evaluation of this report is very mixed, but we do note several notable accomplishments with this study. First, the authors have reviewed and synthesized a tremendous amount of information from the watershed. More importantly, the land use/ USGS loading study provide a considerable amount of new information (land use, loading patterns) which has diverse uses to our efforts to reduce nutrient loading to the bay. Lastly, development and review of the eutrophication index suggests some previously unrecognized patterns and trends in the ecology and condition of the bay which merit close scrutiny in our efforts to better understand and protect the bay.

Nonetheless, we have considerable reservations about this study in its present format. The overall report is poorly and inconsistently organized; methods are inconsistently and, in some cases, inadequately explained. In particular, the section on the validation of the eutrophication model does not lay out the appropriate methods or results (the fit of the 2011 data to the eutrophication model). The manuscript is oftentimes redundant (e.g., see "Statement of the problem and Scale of Ecosystem change). I find statements throughout the report unsupported by evidence or citations.

I am troubled by the big picture portrayal (i.e., insidious system-wide decline due to nutrients) when different bay segments have different patterns and trends (and may even be improving?) according to the available data. The authors may be right but, if so, then our understanding of the bay's ecology is lacking and/or missing some sorely needed pieces.

The water quality data collected as part of this project should be provided including, for each sample, value for each parameter, date and time collected, and GPS location (latitude and longitude), and depth, along with a map of the sample locations.

Comments on Report Organization

The organization of the report is awkward and redundant while missing critical details, which made review challenging.

Some concerns about this report arise from the task of integrating so much information; this requires that the organization of the sections and the information within each section be as consistent and concise as possible. Many sections are redundant and repeat material from previous sections. Wherever possible, redundancy should be eliminated.

The title of the manuscript begins "Nutrient loading" but such information is provided as an appendix and is not well integrated into most sections of the manuscript. The USGS component appears generally well written and should be among the first components of the study that is presented and discussed. The landscape information used to develop the loading is important and useful; these components merit identification as central components to this study.

Comments on Methodology

Data from a separate survey were used as part of the project but no details were given on how these were integrated, particularly since the seagrass species are different between the sections of the Bay.

USGS and Rutgers raw data should be supplied to EPA and NJDEP. This should include GPS coordinates and QC information.

The QAPP (pg. 58) stated that remote sensing data were going to be used as part of this project to assess bay-wide seagrass distribution but there does not appear to be discussion of whether this was done or what the results were.

General Question on how the raw scores are determined: Are the raw scores determined by observed data input into an equation and then turned into a dimensionless unit? Or are the raw scores not all observed data but picked on best professional judgment, literature, and observed values?

Unfortunately, the methods are inconsistently organized among sections and poorly described within some sections. This is particularly true in the index development and the model validation section. The report should more simply and concisely present the methods, including the data collection of data, statistical tests used to analyze the data, and identify one or more appropriate citations regarding their use.

The purpose for including some data and/or methods in the project is not sufficiently clear, even in cases where I think we are in agreement regarding the results and their interpretation. For example, we have many measures of eelgrass condition included in the eutrophication index. What is the benefit of incorporating and analyzing so many different metrics for eelgrass? The utility of the weighting of index scores is unclear; what was the benefit of using weighted and/or final scores? It is not clear what was done in component 4, that is, it is not clear how the model was validated using 2011 data. Component 4 simply presents the results of the data collected during 2011. It should use the model to make predictions of the biological information from the other noted conditions and assess the match between observations and predictions.

Lastly, the manuscript veers off to address other issues (e.g., NCA sampling design) when discussing methods. The "adequacy" of data depends on their use; criticisms directed at other uses of any data are irrelevant and should not be included in the manuscript.

*Thresholds are not clearly defined. As a result the development of raw and weighted scores as well as the overall index of eutrophication can't be properly evaluated:*The selection of thresholds to be used in the assessment of eutrophication in Barnegat Bay was a topic of discussion at the March 28, 2012 meeting. Similar comments were raised in the comment letter on the April 15, 2012 quarterly progress report.

Rutgers provided a response to explain what information would be considered in establishing the thresholds, however, it remains unclear what values were actually used and why.

The original expectation was to identify thresholds that could be applied to other coastal waters. The selection of thresholds must be clearly documented and transparent. Any limitations on the applicability to other waters should be noted. The report includes several references that were consulted. See: Ecosystem State: Water Quality (page 66), include the following statements:

- Kemp et al. (2004) list statistically derived concentrations of dissolved inorganic nitrogen (DIN) and dissolved inorganic phosphorus (DIP) beyond which submerged aquatic vegetation is not present at a variety of salinity regimes. See Table 3-3.
- Wazniak et al. (2007) summarized pertinent thresholds regarding dissolved oxygen, (see Table 3-4) and for total nitrogen, total phosphorus, and chlorophyll *a* (see Table 3-5)) for Maryland's coastal bays.
- For BB-LEH, dissolved oxygen thresholds were defined relative to the New Jersey standard of impairment, which is established at 4 mg L^{-1}

Kemp derived DIN and DIP concentrations of 0.15 TN and 0.01 TP which are expressed as median values calculated over the growing season, whereas Wazniak values for TN and TP are annual averages of 0.55 TN and 0.037 TP. Figures 2-1 thru 2-7 present the maximum, minimum and average concentrations by year. However, based on the threshold selected the data should be averaged appropriately and presented. Figure 5-2 presents nitrogen concentration by seasons. This figure should be revised to be consistent with the threshold selected. If Kemp is the basis, then the figure should present median concentrations during the growing season. If Wazniak is the basis, then annual averages should be used.

The adopted water quality criteria for dissolved oxygen were used even though Table 3-4 includes criteria that might be more appropriate for Barnegat Bay. No discussion is provided to establish what indicates the adopted water quality criteria are the appropriate water quality criteria. The DO criteria developed for seagrasses for Maryland Inland Bays were included but ignored.

The thresholds selected for Light Availability metrics are also unclear. Figure 3-20 identifies several studies with differing maximal depth limits, light attenuation coefficients, and minimal light requirements. Associated light attenuation by various factors such as plankton (chlorophyll a), total suspended solids, and macroalgae are indicated as Kennish et al. 2011 and provided in Table 3-7 but this table describes Area Normalized occurrence of macroalgae. It is necessary to explicitly define which thresholds were selected and on what basis.

Component selection incomplete and duplicative; threshold basis not documented; index derivation has weaknesses: The researcher selected six components to use in describing the biotic condition. There is no definition of "eutrophic condition" so the term should be replaced with "biotic condition" as a less cause-based term. As defined, the components consider factors that overlap across component types, which would tend to afford greater importance to those factors. For example, watershed pressure considers TN and TP loading, while water quality considers TN and TP concentration. Light availability factors are dependent to some extent on the water quality factors. To partially alleviate this effect, in the derivation of the overall eutrophication index, the watershed pressure index is not included to avoid conflation of independent and dependent variables. However, other factors, like physical stressors are not considered in the index at all.

These other factors may be difficult or impossible to consider at this time, perhaps due to lack of data or lack of complete understanding regarding thresholds of effect. They may include community make up of

algae, effect of entrainment/impingement in the Oyster Creek cooling water system and from extensive watercraft traffic, and the effects of resource harvest (has it been/is it at sustainable levels?). Only presence or absence of HABs are considered in the index, even though it is acknowledged that greater dominance of pico algal forms is likely to negatively impact clams. Failure to consider other relevant stressors increases the uncertainty of the findings and these uncertainties should be discussed. As stated, design of an effective management response depends on identifying the suite of significant stressors and determining the relative importance and role in cause and effect relationships with the biotic response indicators of concern.

To develop the index, the study relies on existing data supplemented for some factors by new data collection. Raw data for each factor are provided, not including the more recent data collected through the intensive Barnegat Bay study, which was beyond the SOW. For many years within the span of the study, there is little or no data on clams and benthic organisms. SAV data are sporadic. Only water quality data are generally available. Thresholds are selected for each factor to associate with assessments of highly degraded, poor, moderate, good, excellent. Given the methodology used to derive a unit-less score for the index, the index assessment for any given year is opportunistic (limited by the data available for a given year) and not deterministic (informed by data from the full suite of prospective relevant factors). As a result, the importance of setting thresholds against which observations are compared to determine the assessment cannot be overstated. As the value for each threshold is one of the most important elements in determining the outcome of applying the index, it is essential that the threshold values be solidly based in science. A key requirement expressed as part of earlier reviews of interim products was that the specific basis for selecting thresholds should be identified and substantiated. Instead, only a generic description of the types of sources used for selecting thresholds was provided. This deficiency must be addressed to allow a complete assessment of the tool.

Determination of index values blends raw scores (comparison of average of raw data to a selected threshold) and weighted scores (square of eigenvector value, considering the factors for which there was data in a given year). Weighted scores simply represent a measure of the variability of the factor, if it is present within a given year. If there is no data, the factor is given no weight. The purpose of blending the weighted score with the raw score is unclear. Supporting basis for the statistical approach is needed. Again, because the index score depends entirely on the selected threshold, the basis for selecting thresholds must be thoroughly documented and justified. For the overall index, depending on data availability for various factors, the index may be derived from one or two factors. This is particularly problematic because, as previously stated, there has been no demonstration that the factors selected are the only (or even the most important) factors responsible for the biotic condition.

Comments on the Statistical Approach

- The NOAA ASSETS approach that was specified in the QAPP was not followed. It was fine that an additional approach, the use of PCA, was evaluated, but the original approach also should have been pursued or an amendment to the QAPP sought. The TAC has been continuously asking for additional information as to whether the PCA is appropriate in this case and we find this report still does not provide the necessary detail to evaluate its use.
- The treatment of non-detects has not been adequately discussed. On page 232 it is mentioned that phosphorous has a high proportion of non-detects. This indicates that it is an issue and needs to be addressed.
- The summary tables present minimal information. Since there are non-detects, skewness, and possibly other issues of concern, more thorough summary tables should be presented. They should include the number of non-detects and information on the distribution of the detection limits. Some comments on outliers and skewness should also be included in the table discussion.

- On page 67 a mention is made of zero values in the data. Are these non-detects? The treatment of non-detects needs to be discussed for all analyses in this report, including PCA, Regression, Correlation and ANOVA.
- The approach taken in using PCA (pg. 18) in this report is not standard and no documentation is presented to justify it. Typically, to develop an index using PCA, the scores of the first few principal components would be examined. If the first eigenvalue (score variance) comprises a large amount of the total variability, then the first principal component might be taken as the index. If weighting the index is desired then the first eigenvalue would be used as a weight. In this report, there do not seem to be any attempts to assess the adequacy of using only the first principal component.
- The approach taken in this report is to use the squared component of the eigenvector as a multiplicative weight for that component of the index. The justification is that this weight would be the variance of the component. This claim is not correct. If the variables had been standardized to a variance of 1, then there would be some basis for this, although correlations between variables would also have to be considered. The SAS code in the appendices shows that no variance standardization was done during the PCA analysis and it did not appear to have been done before that. The use of multiplicative weighting should be justified as well as this particular weighting method.
- The only justification for combining the weighted and raw indices is given on page 78. This states that it integrates the multiple indicators and their variability. The advantage of this approach is not obvious and requires some justification and documentation. Combining the two indices might serve to blur any useful measure rather than improve it.
- The PCAs were performed on one to three variables (pg. 369). A PCA on one variable provides no information and should not be included.
- The variables have large coefficients of variation, indicating skewness or some large outliers. Since the eigenvectors in a PCA will be influenced by this, the skewness and outliers should be studied and possibly remedied by transformations.
- On page 185 some regressions are discussed. They all have very low R^2. Even though some of these are significant, it would call for a careful examination of the data to determine whether the regression relationship is valid. Residual plots should be presented to look for deviations from a linear fit.

Comments on Interpretation of Results and Conclusions Drawn

There are no key findings related to the USGS part of the report including the fact that lawn turf represented about 25% of the developed land area and represent a significant source of nutrients to the bay. The USGS data is an important study on its own, as it demonstrates the spatial and temporal variation of nutrient loads to north, central and southern regions of the Barnegat Bay – Manahawkin Bay - Little Egg Harbor estuary. This information is extremely important for developing watershed specific nutrient management plans for the future.

Statements are made throughout the report that seem to represent opinion and are not supported with citations. For example, page 24 and 96 both state “Point source impacts of the OCNGS…significantly increase mortality of estuarine and marine organisms that inhabit the estuary”. Page 26 states “many areas in the north segment of the estuary are now effectively non-swimmable due to sea nettle occurrence”. Page 94 states that a macroalgal bloom in Seawood Harbor caused respiratory problems in residents of the area – were these problems documented or quantified? Unless these statements have supporting scientific documentation, they should be removed.

The new land use and nutrient loading information generated for this report appear solid; we must make better use of them in developing the overall conclusions of this study. Taken together, the two pieces (land use and loads) provide a strong argument that the nutrients in the system are from the developed landscape and that more actions must be taken to address nutrient delivery from the developed landscape.

However, there are some troubling inconsistencies in this report. Perhaps most alarming, the bay is reported to exhibit "an insidious system-wide decline" yet the study's findings point out that the different segments and components of the system exhibit distinct patterns and trends, some of which may even be improving. How do we resolve this inconsistency? First, we must make certain that we identify the weaknesses of the study, which makes use of many different data.

Do we have a handle of the drivers of the bay's eutrophication? Contrary to the author's claim, Guo and Psuty 2000 and Guo et al 2004 suggest that the nearshore coastal ocean, which receives considerable secondary sewage effluent discharges, may be serving as a considerable source of nutrients to the system. The study's conclusions must more clearly identify the weaknesses of the study, which relied primarily on other sources of data. Are we comfortable with the assessment of the eutrophication in the middle bay, which has less developed landscape, is close to the Barnegat Inlet and thus should have better flushing (or is it getting an unrecognized influx of nutrients?), and an "unassessed" potential driver of eutrophication (OCNGS). The authors define eutrophication slightly differently than others; this possibly confuses readers regarding study approaches, causes and effects.

The report includes too many statements which appear to be opinion or are not supported adequately by citations or presented evidence. The management section does not adequately recognize the weaknesses in the project (the inconsistent data from many sources and other unknown components), and should have made better use of the USGS component to more emphatically indicate the need for corrective land use actions. I would also appreciate better attention to citing other sources appropriately.

The development of the biotic index part of this study should be described and the limitations to the outputs and scores listed. Ultimately this part of the project was a modeling study meant to make more quantitative a subjective NOAA model (ASSETS). The results presented exclude the serious caveats listed in the body of the report as to the Index's poor predictability at times due to uneven datasets (some years have no data for some variables) and how these affect model outputs. The researchers themselves state this reduces the applicability of the index. The more important point in the Key Findings section should be the accuracy and predictability of the model.

Report heavily weighted toward SAV; conclusions not in line with data: The report acknowledges that multiple years of condition data for various indicators of biotic condition are lacking. Only one year of benthic data were deemed useable. Clam data are limited. Even though data for SAV are not available prior to 2004 and widgeon grass data were first gathered during this study, the focus of the report is on the condition of SAV, with some discussion of clams. Declines in both SAV and clams are attributed to nutrient loading, in that nutrient loading affects light availability and algal community structure. It is stated that both resources have declined absolutely and also correlatively with an increase in nutrient loads. While absolute measures of SAV biomass and numbers of clams support that these two resources have experienced decline, the unconditional causation statement regarding nutrients is not well supported by data in the Bay and the literature.

Literature supports that light availability is an important factor regulating seagrass growth and distribution. But light availability is affected by both factors that respond to nutrient enrichment (algae in the water column and epiphytic, macroalgal shading) and suspended solids, which consist of both biotic and abiotic components (algae, sand and other inorganic particles). Further, light availability is not the only important factor regulating SAV success. Factors besides light availability noted in the literature

include: physical exposure, substratum, carbon, nutrients, temperature, salinity, oxygen, sulphide, competition and grazing. An obvious example of the importance of other factors is salinity, which limits the distribution of eelgrass in northern Barnegat Bay; in the northern Bay, widgeon grass is the dominant SAV species, as it prefers less salty water. These other factors should be explicit in the report.

In addition to preferring more saline conditions, *Z. marina*, 1) prefers colder temperatures, with a range of -1° C in winter and 25° C in summer and 2) provided there is sufficient light for the species to succeed, at the lower end of the availability range, blade length increases and shoot density decreases, whereas with greater light availability, blade length decreases and shoot density increases. The study indicates the Barnegat Bay mean temperatures appear to be relatively steady, but monthly means as well as the maxima do exceed the 25° C threshold and appear to be trending upward in the north and central parts of the Bay. This suggests that temperature may be a significant stressor for eelgrass. An explanation is required for the selection of the temperature water quality threshold which produces a favorable temperature index value, with an improving trend, contradicting this literature information. Further, with the exception of macroalgal shading, light availability factors appear to be favorable in the Bay, challenging the notion that light availability reductions are a significant cause of eelgrass biomass declines. The exception is the index value for macroalgal shading, but there are only a few data points for this factor. An explanation is needed for how the threshold was selected for the index. While overall biomass has declined, it is noteworthy that the *Z. marina* in the Bay exhibits shorter blade lengths and higher shoot density currently compared to the beginning of the study period, the same pattern as would be expected with a favorable light availability (assuming it is within the required range in the first place), according to the literature.

Another problem with attributing SAV biomass decline solely to nutrients is that total nitrogen levels appear to be stable over the study period. On the other hand, temperature in the bay appears to approach and exceed the physiological limits for eelgrass survival. Physical threats, such as from use of watercraft in SAV areas, were not discussed at all. This would suggest that factors other than nutrients must be considered in determining if 1) SAV can be restored given the temperature regime and 2) if SAV are still viable in the Bay, what are the most important stressors to address to improve the condition of the SAV resource.

The discussion on Biotic Response: Benthic Macroinvertebrates seems to indicate that several indices were considered but the REMAP index was used. The sources for other indices listed are not included as references. Please clarify what validation was done. See page 68 - "Validation of the methodology is conducted through comparison of multiple similar methods, and the response in 2011, as data from that year have been kept separate and out of analyses thus far."

Suggests strong correlations and trends not supported by data and statistical tests; index conclusions unsound: In the description of several factors, a negative trend of biotic factors over time is suggested, but then a variable response is described year by year and an overall conclusion is reached that the "decline" tracks with nutrient increases. Yet the water quality factors remain relatively constant according to the data presented. Under these circumstances, it is difficult to accept a cause/effect relationship between the two. Examples: macroalgae coverage is higher in 2004, lower in 2006, higher in 2008 and lower in 2010, with the 2010 levels lower than the 2004 levels; eelgrass biomass has gone down-up-down-up as well, with the 2010 levels lower than the 2004 levels; water quality parameters have varied slightly but are not largely different at the end of the study period compared to the beginning—except that TP has increased. A number of correlation comparisons are offered in the report along with a conclusion that there is a correlation. However, not surprisingly, based on the above examples, the r or R^2 value provided suggests little to no correlation. This uncertainty is not highlighted in the findings or the executive summary.

Even with the limitations already expressed with respect to developing the index, the researcher describes watershed pressure as moderate to good in Central and South, highly degraded in North; water quality as moderate or good in all three segments, except poor in 2010; light availability as moderate to excellent in North and South, but highly degraded Central; and seagrass response as highly degraded or poor in Central and South (eelgrass not present in North). So, using the index created in the report, it would seem difficult to conclude that water quality was the primary causal factor responsible for the condition of seagrass (data limitations here are that eelgrass data are absent prior to 2004 and widgeon grass was only quantified in this study; no earlier comparisons are possible). The overall index is given as moderate or better in Central and South, but poor in the North (where light availability is generally high, except for a dip in Secchi depth in 2008). With the inconsistencies between the biotic response and water quality, as well as the limitations noted (significant gaps in data), the findings must include caveats as to the uncertainties.

With only a few years of data it is hard to assess (and threshold value basis unknown), macroalgal coverage index is widely variable and very low in 2009. This would be a response indicator likely linked to nutrients and light availability, so warrants deeper investigation re: validity of the index threshold values as this could have a bearing on SAV, benthic organisms and DO. This data gap and its concomitant influences or uncertainties must be addressed.

Estuarine segmentation – in many places the report indicates that the north, central and south sections of BB-LEH behave differently. However, metrics, thresholds, data and conclusions do not appear to be based on segmenting the bay.

Quarterly progress report for the period January 1, 2012 thru March 31, 2012 included graphs showing data collected for water quality. Total Nitrogen was less than 800 ug/l for all years in all three segments. When the total nitrogen scores were presented on a scale of 0 to 100, the scores for the central and south segments were generally 80 to 100 or “excellent”. The north section dipped to about 70 or “good”. While there appears to be a little more variability in the north section, the south and central are relatively stable. This does not support the key finding – “the condition of BB-LEH has progressively worsened over time for both nitrogen and phosphorus. The basis for the statement must be explicit.

It appears that several of the conclusions drawn are not supported based on the historical data provided in the Ecology of Barnegat Bay New Jersey, Kennish M.J. and R.A. Lutz 1984.

Macroalgae: Recurring macro algae blooms *Ulva lactuca, Gracilaria tikvahiae, and Enteromorpha intestinalis,* are listed as a cause impacting the growth of eelgrass. See page 87. From 1969 to 1973 *Ulva lactuca* was the most dominant macrophyte species present in the Barnegat Bay with some of the others listed being in the top ten. Is the current condition worse than what was documented in the late 60’s and early 70’s?

Phytoplankton: A portion of the NEIWPCC Report talks about the change of the phytoplankton to smaller species or picoplankton, which negatively impact the clam growth and is a sign of ecosystem change. What is not discussed is the large abundance of small plankton species present during the summer months (June thru Sept) that were called ultraplankton, very small plankton species, present in the Bay up to 800,000 cells/ml, during the late 60’s and early 70’s by Kent Mountford. See page 64. The presence of picoplankton is not new in Barnegat Bay.

Chlorophyll a: Chlorophyll a range in Barnegat Bay is generally 1-12 ug/l, with a maximum concentration in excess of 40 ug/l between 1989 and 2010. Samples collected in the central portion of Barnegat Bay in a 22 month period from 1969 to 1970 show generally the same general concentration of

1-12 ug/l with a maximum of 35 ug/l. See chart on page 69. It would appear that chlorophyll a levels have not changed much over the past forty years.

Hard Clams (*mercenaria mercenaria*): The Division of Fish and Wildlife studies have documented a population decline and poor recruitment in Hard Clams in the Little Egg Harbor. However, the hard clam population in the central portion of Barnegat Bay was determined to be less than 0.1 clams/ sq. meter in the 1960's. See page 178 -180. It appears that the historical levels were less than 0.7 clams/ sq. meter, the minimum density suggested to be necessary to sustain a viable population.

Nutrients: Nutrients (Total Nitrogen and Total Organic Nitrogen) in the Manahawkin Bay to Little Egg Harbor area seemed to have decreased from the 1970's to the current 2011-2012 data, even though these areas have seen some of the largest percentage change in population growth over the last 40 years. Stafford Township, 720% increase, Little Egg Harbor, 675%. Stafford grew from about 4,000 in 1970 to approximately 27,000 in 2010. Little Egg Harbor grew from 2,972 in 1970 to 20,065 in 2010. Data from research performed by Durand from Manahawkin Bay to Little Egg Harbor, in the book by Kennish M.J. and R.A. Lutz 1984, Ecology of Barnegat Bay New Jersey, and comparing it to data collected during the Barnegat Bay Sampling from the summer of 2011 to the summer of 2012. In spite of the population increase over the last 40 years, Total Nitrogen (TN) and Total Organic Nitrogen (TON) concentrations in Manahawkin Bay have declined from an average of 510 µg/l TN and 489 µg/l TON, while the 2011-2012 data shows 340 µg/l TN and 291 µg/l TON. Reviewers would be interested in knowing the author's reasoning for the divergence of his historical data and current conclusions.

Page-Specific Comments

Acknoweldgements:

- The acknowledgements section should include all members of the technical advisory committee. Debra Hammond, Stan Hales, and Jim Vasslides have been left off the list.

Key Findings:

- Includes an extensive use of acronyms that make this portion of the document difficult to read. It may be helpful to provide the full terms and information that highlights the significance of the findings for the general audience in this section.
- Having a key findings section is a good idea, but the section needs to be simplified. As is, this section is not reader friendly for a general audience.

Page 13:

- First bullet - This is not a finding but a statement of condition. Loads need context (how high is this comparably) and the detailed USGS modeling info on both nitrogen and phosphorus, as well as the differences in spatial loadings from north to south.
- Second bullet -No context is provided. There are no indications of how frequent or what percentage of the measurements were exceedances for many variables and any trends analysis performed. An increasing or decreasing trend would be a key finding. The above statement also implies that all of the above referenced factors directly led to degradation of sea grass and led to mortality. This is not a finding of the model nor are some of the variables related to sea grass degradation (e.g. fewer hard clams and food web shifts).
- Third bullet: This needs clarification. Sea grass response "to what" is highly degraded. True, sea grass abundance and vitality were shown by project monitoring to be declining over time in the field studies. However they are but one set of variables among many in the model for assessing nutrient impacts. It's unclear what factors seagrass is responding to. Other stressors not included

in this study are physical and mechanical impacts from boat traffic and bulk heading effects causing the loss of habitat. These variables were not included in the model nor studied in this project.

- Fourth bullet: "Data availability limits its power." What is the source of the variability, mentioned here only in passing, but discussed in much more detail within the body of the report. Until the researchers address these scientific concerns it's unclear whether the model outputs are accurate or significant. Understanding the limitations of the data inputs and model outputs will determine whether the Index of Eutrophication is ready for use with acceptable accuracy for environmental management purposes.
- Fifth bullet: These overall scores and statements about declining eco-health leave out the more detailed discussion of data limitations in text which highlight that due to data paucity of certain variables in different place and years the index may have different scores based only a few of the variables (e.g. there is no eel grass or brown tides in north).
- Sixth bullet: When a model plateaus and variability increases it usually indicates a problem with either data quality, model imbalances (weighting or scaling of variables), or other factors external to the model (stressor data not included in model). This variability should be explored mathematically by evaluating data sets and modeling parameters (equations). This analysis should be done before modeling scores are reported above, as it's unclear if index scores outputs are meaningful.

Page 14:

- First full bullet - Does condition mean index score, or loadings, or specific ecological degradation (SAV, hard clams)? Regardless, this broad conclusion is not supported based on the historical data provided in the "Ecology of Barnegat Bay New Jersey," Kennish M.J. and R.A. Lutz 1984. For example, recurring macro algae blooms are listed as a cause impacting the growth of eelgrass back then. Is the current condition worse than what was documented in the late 60's and early 70's? For phytoplankton, a portion of the Report indicates a change in phytoplankton to smaller species or picoplankton, which negatively impact the clam growth and is a sign of ecosystem change. In the late 60s and early 70s ultraplankton, very small plankton species, were present in the bay at up to 800,000 cells/ml. Thus the present of picoplankton is not new to Barnegat Bay. Between 1989 and 2010 the chlorophyll *a* range in Barnegat Bay was generally 1-12 ug/l, with a maximum concentration in excess of 40 ug/l. Samples collected in 1969 - 1970 showed generally the same general concentration of 1-12 ug/l with a maximum of 35 ug/l. It would appear that chlorophyll a levels have not changed much over the past forty years. Reviewers would be interested in knowing the author's reasoning for the divergence of his historical data and current conclusions.
- Second bullet: How is this statement supported by this model? How is it supported based on the historical data provided in the "Ecology of Barnegat Bay New Jersey," Kennish M.J. and R.A. Lutz 1984? For example, nutrients (Total Nitrogen and Total Organic Nitrogen) in the Manahawkin Bay to Little Egg Harbor area seemed to have decreased from the 1970's to the current 2011-2012 data, even though these areas have seen some of the largest percentage change in population growth over the last 40 years.
- Last bullet: If so, is there causal linkage to biota or are there other more important variables?

Page 16:

- Next to last paragraph on the page, last 5 sentences about the manuscript's organization: Based on these five statements, I expect the other sections of the report to discuss 1) loads, 2) stressors and responses, 3) index development and current condition, 4) eutrophication, and 5) conclusions and management recommendations. I expect the sections to be organized consistently (e.g., intro,

methods, results, etc.). Also, eutrophication extent is identified redundantly as a goal of component 2 and 4.

Page 17:

- The report should discuss how the specific candidate indicators were selected. Were any minimum criteria applied for indicators?

Pages 18 & 61:

- These sections describe the calculation process of the biotic index of eutrophication for the estuary. This process includes calculations of "component index scores" using "raw scores" and "weighted scores" for all 21 indicators, calculated for six components using data collected each year (1989 through 2010 and 2011) in 3 segments of the estuary. This process is complicated, thus a simplified "step by step" diagram/schematic providing a big picture of how the final "overall" index for the estuary was developed would be helpful.

Page 22:

- Last complete paragraph on the page (before INTRODUCTION): I'm not sure what this paragraph is trying to say; it looks like a statement taken from a proposal. As the final paragraph of the Executive Summary, I think this paragraph should provide 1) a summary of the key recommendations for addressing the bay's eutrophication via nutrient management or other actions or 2) the key next step (e.g., address one or two primary data gaps re eutrophication, etc).

Page 23:

- Last sentence in second full paragraph: “Extensive studies conducted on the estuary during the past two decades have documented these problems” is then followed by citations of four sources written by one of the authors of this report (Kennish). I would imagine there are studies by other authors and it would be more credible to include some of those sources as well.

Pages 24-27:

- “Statement of the Problem” and “Scope of Ecosystem Change”: these two sections are redundant; they should be combined and shortened.

Page 26:

- “Since 2004, eutrophy has generally worsened in much of the BB-LEH, and the condition of the seagrass habitat has significantly degraded.” How is eutrophy defined? Increased chlorophyll a, low dissolved oxygen levels, excessive diurnal swings in dissolved oxygen, plant biomass or is the conclusion based on seagrass decline? Please add definitions to the report.

Page 27:

- “Seagrass now covers a 5260-ha area of the BB-LEH estuarine floor (Lathrop and Haag, 2011).” This means that seagrass covered about 20% of the bay bottom. Does this include widgeon grass in the northern part of BB? What percent of BB-LEH could/should support seagrass?
- “Designated as moderately eutrophic in the early 1990s, BB-LEH was later reclassified as highly eutrophic in the late 1990s, a designation reconfirmed in 2007.” Bricker’s approach was very subjective and heavily weighted towards brown tide blooms. The reason for this project was to develop an index appropriate for BB-LEH.

Page 29:

- “Exchange of bay and ocean water occurs through these three inlets. The continuity of the barrier island complex restricts the exchange of water with the coastal ocean, resulting in a protracted

water residence time in the estuary amounting to 74 days in summer when eutrophication is most problematic (Guo et al., 1997, 2004)." Barnegat Bay functions as three separate waterbodies and would be expected to have different retention times. How does this factor into the overall eutrophication index for the North, Central and South?

Page 30:

- "Nonpoint source inputs account for almost all of the nitrogen entering the estuary." The role of the OCNGS as a "point source" of N has not been evaluated. Chlorine used daily for 2 hours at the plant interacts with organic tissues and forms chloramines, the fates and fluxes of which are presently unknown. Essentially this warm chloramine soup may be considered a "source" of N, as living tissues are being broken down into nitrogenous forms which may affect bacterial, other microbial, and/or phytoplankton production.
- "Only a minimal, diluted fraction of this effluent may re-enter the estuary via inlet exchange and would not be considered a point source of nitrogen." Guo and Psuty 2000 suggests substantial N may enter the estuary through inlets based on inlet N measurements. Guo et al. 204 makes additional arguments in support of the ocean as a source of N from offshore sewage discharges from an examination of current vector diagrams.
- "Confined animal feeding operations (52 total) cover a very small area of the watershed. With only one exception of a centrally located feeding operation in the watershed, all are located in the northern portion of the watershed." Citation?

Page 31:

- Study Area Characteristics: Recommend including a table describing the physical and chemical characteristics of each segment of the estuary (e.g., include watershed and bay size, mean depth, sediment grain size, salinity, DO, mean P, mean N, etc). This should be done using the most recent data. Additional columns showing these characteristics at points in the past (where available) would be most useful.
- Water Quality: indicates several bases for high nutrients in the northern section. However, figure 5-2 seems to conflict with higher levels observed in the central section.
- Section re OCNGS: I don't know what the greatest impacts of the OCNGS are, because, as noted in the manuscript, the OCNGS effects have been so inconsistently assessed. Everyone talks about the OCNGS impacts from entrainment and impingement on the affected populations, but the ecosystem impacts (and not just the population impacts to impinged/entrained spp) also merit scrutiny. The OCNGS is acting as a predator and "cropping" the production of long-lived life forms (fishes and macroinverts) in favor of shorter lived species (bacteria, etc.). Shouldn't there be a more holistic assessment of the OCNGS effects?
- "Other adverse effects on estuarine water quality include nonpoint source inputs of pathogens and other pollutants as well as bulkheading, dredging, and lagoon construction." Were any other pollutants or historical pollutant sources reviewed or assessed as stressors? FYI, I think NJDEP data show some EDCs have been sampled in higher concentrations in the BB than in any other system in NJ.

Page 33:

- Sections in this report are not consistently organized.
- A large portion of Objective 4 was removed (as compared to the QAPP). Left out was: "Each of these (estuarine biotic responses) will be examined and assessed for statistical validity and inclusion in the index development for the 1989 to 2011 period."
- Objective 5 was changed from what appeared in the QAPP, now says: "to generate..." instead of "to develop" a biotic index.

Page 34:

- "We have attempted to characterize the spatial and temporal dynamics of the nutrients within the estuarine system that could be used to establish the basis for developing accurate nutrient loading criteria. Based on these findings, we have modeled how estuarine health will likely change as a result of several important policies for land use and nutrient pollution control.**"** Elsewhere in the report, the limitations on the data are documented. Using the existing data set, it is unlikely than any attempt to develop nutrient loading criteria, along with the appropriate errors bars, would be technically justified. The data do not clearly support the suggestion that, if we reduce the nutrient loading to a specific level, the seagrass beds will rebound, hard clam recruit and populations will improve and sea nettles will disappear.

Page 35:

- Watershed Nutrient Loading. The QAPP indicates that only secondary data collected by the Bureau of Marine Water Monitoring for the period 1989 thru 2010 would be considered. However, this section implies that data collected by the USGS and available thru NWIS were also considered. Did the USGS work look at both DEP and USGS data?
- Component 1, "Watershed Nutrient Loading Loading." Note the text duplication.
- "After thoroughly reviewing aspects of the data such as units, detection limits, and site locations, a database of quality-assured water-quality data was developed. Additional information is needed here so we clearly understand the basis for selecting data. The reasons for the inclusion and exclusion of data must be more clearly and simply described. It is essential. Either here or on page 61 under "Data Gaps", further detail needs to be provided what the criteria were for this review, specifically what temporal and spatial coverage criteria were used, if and how outliers were determined, how missing data were treated, what the criteria were for representativeness and age of data, sampling and laboratory method comparability, etc.
- What level precipitation and hydrologic data were used to segregate base flow from runoff?
- Last paragraph. What is a baseflow separation analysis: is it the identification of data collected during baseflow conditions?

Pages 35-39:

- The report states: "A numeric scoring system was used that computes an index value from key water quality and biotic indicator measurements in each of the three estuary segments for years sampled during the 1989 to 2011 period". This statement is misleading if the index scores for some of the components were calculated based on data collected during only one year period (e.g., benthic invertebrates). In the discussion of components 1 through 5, it becomes clear that different datasets were used to evaluate different biotic responses. (e.g., data on seagrass beds were collected from 2004-2006 and from 2008-2010, data on epiphytes were collected from 2009-2011, while hard clam and benthic invertebrates data were collected in 2001). On page 140, Table 3-2 nicely summarizes all datasets used for evaluation of different indicators. It would be helpful to refer to this table on page 35.

Page 36:

- Estuarine Biotic Responses: The final report should include the citations of field and lab methods and QA/QC and any modifications that were made. It should also cite the QAPP.
- Estuarine Biotic Responses: Secondary water quality data were used in this investigation. Describe how these data were screened for use. In particular, what temporal and spatial coverage criteria were used, if and how outliers were determined, what the criteria were for age of data, comparable sampling and laboratory method criteria, etc. Representativeness of the secondary data also should have been evaluated and needs to be stated. Some of the secondary was

collected in a targeted or focused manner, i.e. to find problems, and is not appropriately used in an index that is determining overall condition. This information should be summarized in a table along with source citations for the data.

- Shellfish data were collected at regular bi-monthly intervals from June to November. However the text indicates that the Department collected data from July 16 thru August 31, 2001. Please clarify or correct.

Page 37:

- Phytoplankton section describes the Department's remote sensing program which started in 2011. How were the data used? Were only water samples collected in the blooms used? Also, other reports published by DEP and USEPA indicated other blooms of green and red tides occur periodically and in other parts of the bay beside LEH. Why wasn't this data set used as well in HAB part of index?
- Light Availability. Several metrics are listed here. However, the researchers indicated that the percent of light available to seagrass leaves was a better indicator than secchi depth and includes chlorophyll a, total suspended solids, macroalgae coverage, and epiphyte coverage. Since these are factored into the calculation, incorporating separate measures for these parameters is essentially double counting. Please explain the rationale for the approach used.

Page 38:

- Benthic Macroinvertebrates: an expected outcome of this project was the development of a benthic indicator that could be used to assess current conditions and changes over time. It does not appear that a stand-alone benthic index has been considered. Benthic Invertebrate Response is blank on Table 3-2.
- Benthic invertebrates. Citations are needed here. I would avoid use of the word "balanced" because it is often hotly debated in community ecology.
- "The development of a biotic index includes a benthic invertebrate component, which is needed to measure the overall ecological condition of the estuary." Additional detail is needed about the "index development." How was the index developed? How is it determined?
- "Methodology Elements, Component 3" subsection states that: "For ecosystem pressures, the metrics include total nitrogen loading, and total phosphorus loading". This is not consistent with a statement made in "Ecosystem Pressures" section on page 43, which states "Water residence time and total nitrogen loading are the two key indicators of ecosystem pressure used in this project." On page 39, the report also states that "Such an index would combine ecosystem pressures (nutrient loading and water residence time), ecosystem state, and biotic responses. Please clarify.

Pages 38 and 55:

- A benthic invertebrate response is used as one of six major components used for the eutrophic index development. The report states that benthic invertebrate data collected only in 2001 was used in the development of the eutrophic index for the estuary. Please reference where the 2001 benthic macroinvertebrate data were obtained from. The report states that too few samples were collected in 2000, 2003, 2005, and 2006 to be used for adequate statistical analysis. Could this data be used to validate index developed based on data collected in 2001? Also, were benthic invertebrate data also collected in 2011?

Page 39:

- Biotic Index Development: "An important goal of this project is to develop an effective and useful index of eutrophic condition for the BB-LEH estuary." The actual objective (stated in the QAPP) was: "To develop a biotic index of estuarine condition using water quality and biotic indicators to assess eutrophication, impairment, and overall ecosystem health of the BB-LEH

estuary…." This section only seems to cover how the index was applied to the segments, so a discussion should be included of the application to the entire system. Can the index as currently developed be used to assess overall ecosystem health?

- Biotic Index Development: The QAPP (pg. 28) stated: "These major groups (estuarine organisms) will be monitored across the study period to determine when numeric shifts occur in abundance, …., which will be correlated with nutrient loading levels to document the threshold points and levels of biotic decline." How were these threshold points determined and where in the report are they shown?
- Biotic Index Development: The QAPP (pg. 28) stated: "They (biotic data) will be examined and assessed for statistical validity and inclusion in the index development for the 1989 to 2011 period." The actual statistical procedures and outcomes should be included in the report.
- Validation Dataset (2011): Describe the validation procedures. What statistical comparisons were used for validation?

Pages 39, 55, 60 and 67:

- "Gleason's D value" is used as one of matrixes for benthic invertebrate response component. Term "Gleason's D value" is not defined in the document.

Page 39:

- "Such an index would combine ecosystem pressures (nutrient loading and water residence time), ecosystem state, and biotic responses." What is the basis for using water residence time as an ecosystem pressure? Also, if you are going to examine data by segment, shouldn't the water residence time be examined by segment? Residence time varies among segments; wouldn't it also be affected by the tides (spring vs neap, local sea levels which change with storms (hurricanes, northeasters), and OCNGS operation, which varies seasonally?
- Component 4: Validation Dataset (2011) for Eutrophication Assessment. The objective identified here is different from the statement in the Executive Summary.

Page 40:

- "Component 5: Synthesis and Management Recommendations. This section needs a clearer statement of the approach and methods. To make some of the conclusions made in this section, the authors must also rule out or eliminate some other potential causes (e.g., climate change, contaminants, overfishing) of the changes observed in the ecosystem. Also, there must be a recognition of the weaknesses of the study (data from other sources, lots of data gaps…) The authors may be right, but they need to demonstrate the robustness of their findings and conclusions.

Page 41:

- COMPONENT 1: NUTRIENT LOADING ANALYSIS. As written, this section presents no findings but simply references the USGS report. Organizationally this is extremely awkward, because we have not read this section and now must simply accept information moved from the appendix to the "bigger" study without knowing the approach, methods, etc. In addition, the data generated in this appendix" are invaluable to the report and must be put to better use in different sections of the report.
- The report states that full details on baseflow loads calculations are compiled in Appendix 1-1". Appendix 1-1 consists of over one hundred pages. Please identify a specific page as a reference for baseflow load calculations.

Page 42:

- Estuarine Biotic Responses: The QAPP states that shellfish (bay scallop) data will be collected while in the field. Where are these data in the report?
- Estuarine Biotic Responses: Biotic Response Sampling: A summary of the QA/QC results should be given based on the Measurement Quality Objectives stated in the QAPP.
- Regarding the sampling methods of Short, please clarify statements such as “modification with perpendicular transects”.

Page 43:

- Ecosystem Pressures. I am uncertain why some variables are ecosystem pressures and others are ecosystem states. I suppose a variable can be both; however, this is confusing. Also, this structure is inconsistent with the authors’ unusual use of the word definition (see my definition comment in component 5). If nutrient loads are eutrophication (by the authors’ definition), eutrophication is causing eutrophication. This circularity of logic must be corrected.
- Should precipitation, temperature, and/or turbidity be a pressure given they affect "alternative states." Could the number of boats registered within the water be a pressure?
- Ecosystem State – Water Quality: The summary of current conditions should be consistent with the thresholds selected (annual average, annual median, growing season average). Based on the review of literature provided, none employ the maximum values.
- Please clarify what is meant by “haphazardly” tossing a quadrat into a sampling system.
- Please clarify how the accuracy of diver estimated percentages were assured.
- The report states "Nutrient loads are presented in Appendix 1-1". Appendix 1-1 consists of over one hundred pages. Please identify a specific page as a reference for determination of nutrient loads.

Page 44:

- Ecosystem State- Light Availability: Many of the measurements were not useable because the secchi disk hit the bay bottom. In addition, on page 37, light availability was a better indicator than secchi depth. Please explain the value of this parameter.
- Condition Variables: Temperature, DO, etc. In these results sections, should there be a general discussion of annual and/or seasonal trends?

Page 45:

- Indicates that Twilley documented 25 cm thick bloom-forming macroalgae. Since his work was conducted in Chesapeake, it might be better to delete “In the nutrient enriched waters of this coastal lagoon” or at least change it to read Chesapeake Bay.
- Macroalgae coverage. It appears that macroalgae sampling was limited to seagrass beds in the south and central segments. As this is a parameter included in the light availability factor, how is this factor addressed in the northern section or other areas without seagrass beds?
- Macroalgae Percent Cover: What percent cover had an initial effect on the seagrass based on the literature? Based on this section, 59% coverage does not impact the SAV/ecosystem. This parameter and percent value ranges need to be examined and linked to literature based values (i.e., on an empirical basis).
- Page 45, Macroalgae Percent Cover. Much of the information presented in the first two paragraphs of this section is out of place and should be deleted; only information relevant to the methods need be included in the methods section. Several different types of methods are used to collect different data. I recommend subheadings to make clear the different types of data which are being collected. Why are the authors using so many different measures of macroalgae? What are so many measures needed for macroalgae?

- The third and fourth paragraphs of the macroalgae section are excellent: they clearly present results in an organized manner. All of the sections should be similarly organized.

Page 46:

- "Benthic macroalgae are powerful drivers of change in water quality and seagrass habitat." Citation? Be careful with your choice of words: correlation does not equate with causality.
- "Macroalgal blooms contributed in part to the decline of seagrass biomass in BB-LEH over the 2004-2010 period (Kennish et al., 2008, 2010, 2011)." This appears to be a concluding statement which has not been established by the presented results. The next three paragraphs do not belong in the results sections.
- Macroalgae coverage - It appears that macroalgae varies significantly from year to year. Based on the level of variability, is it reasonable to conclude a trend?

Page 47:

- Epiphyte Percent Cover. The first four paragraphs of the Epiphyte Percent Cover section are interesting but out of place. The methods need to be located consistently with other sections.

Page 48:

- "Epiphytic areal cover on seagrass leaves was determined by collecting the five longest …" Citation for this method?

Page 49:

- Ecosystem Response, Eelgrass. I don't' think anyone questions the importance of eelgrass in the BB/estuarine ecosystems, but, as with macroalgae, what is the purpose/need for so many measures? What do they tell us? Can they be compared with *Zm* losses in other systems to provide evidence for causality of losses?
- Ecosystem Biotic Response – Eelgrass: This section needs to include basic information about eelgrass growth characteristics in terms of what is 'normal' or 'unstressed'. For instance what is the timeframe (e.g., month, temperature, light intensity) when shoot density is normally highest during the year? The same information is needed for biomass, coverage, blade length, etc. This will help the reader understand the various seasonal data discussed.
- "Results of this project show conclusively that eelgrass condition in BB-LEH has declined substantially through time and that the rate of decline is related to nutrient loading and associated symptoms of eutrophication." While this may be true, I don't think this has been proven at this point in the document. After all, isn't that the point of Component 3?
- Eelgrass- While the research may show a decline in the condition of the seagrass beds from 2004, it does not support the finding that this decline is due to spatial or temporal nutrient loadings. The loading analysis provided indicates that 65% of the loadings originate from the Toms River and Metedeconk Watersheds, which are located in the northern section, while the seagrass beds are located in the south and central sections. Further, the loadings fluctuate based on precipitation. This supports the need for the dynamic water quality/quantity modeling the Department is conducting through a contract with USGS in order to define the relationship between nutrient loading and productivity in the bay.
- Return to previous levels of eelgrass biomass may be difficult to attain (Duarte et al., 2009). Findings from this and other studies may be relevant and appropriate for incorporation in the manuscript, but only results should be presented in this section. Didn't eelgrass recover from wasting disease from throughout large parts of its range?

Page 50:

- Eelgrass - Several metrics are used to evaluate the condition of the seagrass beds, some show improvement over time, while others show decline or no change. Which measures are most important and can they be aligned as to which are affected by nutrients versus other stressors such as water temperature or human disturbance?

Page 52:

- Widgeon Grass - The assessment of widgeon grass was added to ensure that the eutrophication index would work in areas with lower salinities that cannot support eelgrass. It does not appear than this work was factored into the overall assessment. How much additional information would be required to incorporate?
- Harmful Algal Blooms – this section should be rewritten. The statement "not been monitored in the estuary since 2004, and thus no observational HAB monitoring data are available over the past eight years" does not acknowledges the Department's use of remote sensing with follow-up monitoring and species identification.
- Other Biotic Components. Again, methods, results, and discussion within the subheadings are not consistently organized. It's not clear how data were selected or used.

Page 53:

- Harmful Algal Blooms (HABs): "Dissolved organic nitrogen concentrations were not directly linked to the blooms, which may be more closely aligned with the concentrations of dissolved organic nitrogen in the estuary." Sentence does not make sense, both factors mentioned are dissolved organic nitrogen.
- "Dissolved organic nitrogen concentrations were not directly linked to the blooms, which may be more closely aligned with the concentrations of dissolved organic nitrogen in the estuary." Citation?
- Shellfish Section, First sentence. What was the pattern in the number of clammers over this period; is there any trend in the CPUE?

Page 54:

- "The loss of such large numbers of hard clams also appears to reflect a shift or transition in the system away from one of top-down control exerted by filter feeders consuming and regulating phytoplankton populations to one of bottom-up control limited by nutrient inputs." I'm not saying that this is incorrect, but the clam data alone are insufficient justification for this conclusion. There are other components in the system which may compete with phytoplankton for nutrients. Why aren't chl concentrations in the system higher if phytoplankton aren't being grazed?
- Benthic Invertebrates. This section is far too much a diatribe against other large national monitoring programs rather than a clear presentation of why data were or were not used.

Page 55:

- Numerous datasets were excluded. Some discussion should be included here regarding what selection bias could occur by allowing and not allowing data sets

Page 56:

- Component 53, Biotic Index Development. This component section has a different structure than previous component sections. Where are the methods? The results of the development of the index are presented as conclusions and are, without knowing what was done, inappropriate.
- The researcher states that the subject Biotic Index of Eutrophication is the most comprehensive and holistic assessment of BB-LEH conducted to date. In order to assess the ~20 indicators, the index integrates over 74,400 observations among 85 variables. However, the sheer number of

observations, variables and/or indicators used does not validate the index. Data availability is likely a big factor is whether the index reasonably performs. The uncertainty must be discussed, including the effect of data availability.

- "Data availability remains a major limitation to assessment of eutrophication condition for BB-LEH. While an increasing number of indicators are being monitored, aligning data collection through space and time and increasing sampling frequency will greatly improve future assessments." A similar conclusion is presented on page 68. We definitely agree. This is so important it should be listed as a "KEY FINDING."

Page 58:

- On the first bullet, just focusing on the Central section: Light availability is increasing but it's greatly worsened in the central section. Chlorophyll a is good, suspended solids is excellent, epiphyte coverage is excellent and the percent light reaching seagrass is good. The overall conclusion was that light did not penetrate deep enough. On page 75, the water quality index was described as moderate and sometimes good. Watershed pressure indicators are considered good. The conclusion provided on page 75 on the seagrass response index indicates the condition is "highly degraded" to "poor". It appears that the factors don't support the conclusion that water quality, watershed pressures or light availability are responsible for the decline in seagrass.
- National Estuarine Eustrophication Assessment. The NEEA and the following Goals and General Approach Section are redundant and should be cut to focus solely on the methods that were developed and/or used. A revised methods section should be the first items presented within this component.
- "Influencing Factors include Load (nitrogen ratio) and Susceptibility." Please explain your use of the term susceptibility. Is there a relevant citation? I thought the "pressures" identified in a previous section were N-load and residence time.

Page 59:

- "The indicators are organized together into six components: (1) Ecosystem Pressures, (2) Water Quality, (3) Light Availability, (4) Seagrass, (5) Harmful Algal Blooms, and (6) Benthic Invertebrates. Ecosystem pressures does not appear to be used consistently throughout the manuscript. I recommend changing "ecosystem pressures" to nitrogen or nutrient loads.
- Background: Building on the NEEA: Where is the "matrix" that is mentioned in paragraph 2?

Page 61:

- PCA: The appropriateness of the Principal Components Analysis (PCA) and its weaknesses should be discussed and related to its use in the index. One of the assumptions for PCA is that sets of scores should represent a random sample from the population of interest. Was that the case in this application?
- Page 61: Please provide a justification for averaging the six variables together.

Pages 61-63:

- "Available Data and Data Gaps: Please clarify the relative valuation of all these data sets. Explain why some datasets were held to stringent screening criteria (e.g., NCA data) while other datasets were used where location of the sampling was unknown (HAB dataset). It may also be helpful to further expand the Section on Data Availability and Data Gaps to describe the quality assurance and quality control of the data.

Page 62:

- Note that applicability of the index to any given segment depends in part on availability of data within that segment. Does this mean that the eutrophication index can only be applied to areas

with seagrass or that an index cannot be calculated if data on all six variables are not available? Considering the gaps in available data, it would appear that the index could only be calculated for a few years in some locations. Please clarify.

Page 63:

- "Examples include NCA data (2000-2006), residence time, hydrodynamic modeling, GIS layers of seagrass coverage, counts of jellyfish, and several others." All considered data sets and the reasons for their inclusion or exclusion should be listed in Appendix 3-2. Please make certain that you cite appropriate sources (e.g., BBP SOTB report, etc).

Page 64:

- "Raw scores all range from 0 (bad) to 50 (excellent)." The rescaled raw scores for TN loading and TP loading range from 0-100, not 0-50. Why is this? The answer comes on page 76, but it should be discussed way up here so it is not so misleading.
- Determining Thresholds: It needs to be clearly presented, in table format, the identification of the final indicator specific thresholds that were used and what each indicator was based on (literature, BPJ, etc.).

Page 65:

- The report states that thresholds for TN and TP loadings were determined by examining biotic responses to nutrient loading reported in the literature and analysis of PLOAD model. A summary table listing TP and TN loading thresholds obtained from individual literature sources would be helpful.

Page 66:

- How was the threshold for nutrient loadings determined, taking into consideration that the loadings are dependent on precipitation?
- In reference to Figures 3-14 to 3-16, how were thresholds determined from the figures? These figures are extremely difficult to understand. Consider breaking up the segments into different figures and providing a regression analysis that can help us determine the effect that total nitrogen loading has with seagrass, light variables, and water quality. For Figures 3-14 to 3-16 and 3-23 to 3-24, what are the blue lines representing? Is this the threshold value?
- Dissolved oxygen thresholds are defined relative to the New Jersey standard of impairment, which is established at 4 mg/L. There should be a good amount of DO data for the bay, did the observed data correspond to this threshold? Wasn't part of this study to determine the site specific thresholds and not to necessarily take the previous standards and apply them as thresholds?

Pages 66 & 67:

- A summary table of all thresholds obtained from literature for water quality, light availability, HABs, and benthic invertebrates would be helpful along with a respective literature sources.

Page 67:

- The thresholds identified for harmful algal blooms are based on the magnitude of an observed bloom. As identified in these comments, the report includes statements that there has been no monitoring for HABs since 2004 and a bloom in 2010. How is the frequency of events or areal coverage factored into the metric?
- Biotic Response: Seagrass: This section states that seagrass assessments were "adjusted" because of "the uncertainty associated with identifying reference conditions in BB-LEH." Describe how these adjustments were done.

- Biotic Response: Benthic Invertebrates: This section does not state what the "thresholds" or "spectrum of reference conditions" were that were used to evaluate the benthic invertebrate data.
- "Data are analyzed separately for each segment of the bay, because they have been determined to be heterogeneous habitats." If this is true for benthic invertebrates, then should all of the figures lump all of the segments together? Would it be beneficial to provide assessments and analysis for each segment separately?
- Was a statistical analysis or regression analysis done to support the claim that sea grass indicators responded negatively to increases in chlorophyll a (Figure 3-23)?
- In reference to the statement, "Data were analyzed to identify if changes in rates of decline (seagrass) were evident with respect to total nitrogen (Figure 3-16)," The figure is difficult to draw any conclusions based on the scale. Can we change the scale? Can we add a regression analysis to the figure? Should the figures be divided, with each separate figure showing the segments separately?

Page 67-68:

- For biotic response, can we provide a figure with the observed response to an increase in nitrogen load?

Page 68:

- Typo "Principal" Component Analysis
- With limited data for benthic invertebrate components, can we determine that the index will accurately represent the response to the benthic communities with load increases? Do we have a comprehensive analysis on the current benthic community status for each segment?

Page 71:

- Figure 3-31 has Water Quality Indicator Scores for DO as between 10 and 40, which is poor. How are these determined, since observations of indicator scores are compared to thresholds (4mg/L, page 66) then the measurements are rescaled into indicator scores (page 56). Is the available data for DO showing an overwhelming poor indicator score in reference to the DO threshold of the New Jersey Water Quality DO standard of 4 mg/L? Even though on Figure 2-2 DO is generally above 4 mg/L. Also, is Figure 3-31 just the raw indicator score? If so, please label the figure.
- With stable Total Nitrogen Loading scores, from 1989 to 2010, how can we then say, "It is clear throughout the entire system, nutrient loading – both total nitrogen loading and total phosphorus loading – has resulted in substantial degradation and eutrophication of BB – LEH" (page 14)? The scores and this statement appear to contradict.
- Indicator Scores. So are these just the rescaled or "raw" scores? Also, there are some apparent inconsistencies here between the presented information and the interpretation. First, nutrient loads have remained fairly constant over the entire time. Second, am I reading correctly that the loading scores are lower in the northern bay segment?
- Figure 3-30. Why are these listed as "raw scores" in the figure legend while all of the other indicators in subsequent figures are not?

Page 74:

- The index values are presented with the descriptors ranging from excellent to highly degraded. Were any other breaks considered such as eutrophic, mesotrophic, oligotrophic? Is there a breakpoint between acceptable and unacceptable conditions?
- Component Indices: The benthic invertebrate index is missing from this section.

Page 75:

- Weightings: Table 3-10 doesn't show the weightings for the benthic invertebrate indicator.
- Watershed Pressure indicator: Is this the same as the Ecosystem Pressure indicator discussed earlier in the report?

Page 78:

- Discussion: Extensive detail is provided in the discussion section, however the overall summary of the "final index" would be helpful. An expanded qualitative discussion of the uncertainties should also be provided here.

Page 82:

- Validation Dataset (2011): This section reports the data collected in 2011 but should focus on how the validation was done.
- Additional data from a separate project in northern Barnegat Bay were used with this study. Since these were not discussed in the QAPP for the current (NEIWPCC) project, discuss method and data comparability, reference the appropriate QAPP, and describe the process and thresholds for incorporating *Ruppia* into the index.

Page 83:

- The validation year for DO was all above the 4mg/l threshold. Does this provide a good fit to the results obtained from the index?

Pages 85 -86:

- Can we see figures that show the observed seagrass data from each year, including the validation period, so that we can see the negative or increasing trend? The information appears to be in separate tables, and this is difficult in determining the validation of the data when separated.

Pages 87-88:

- Regarding conclusions drawn from the validation period - did the observed decreases in seagrass and shoot densities of eelgrass correspond to increased loading of nitrogen for the 2011 year? Can we see a figure that shows the nitrogen loading per year and the effect it had on seagrass for the validation period? Did the observed results fall into line with the predicted results from the eutrophication index? As part of the validation component of the project, the validation set was to show that the index of eutrophic condition from the data collected through 2010 has been applied using 2011 data for validation. Can we show this in a figure with statistical analysis, that shows that the data for 2004 to 2010 and the eutrophication index results have been validated based on the 2011 data set?

Page 87:

- "Therefore, the aboveground and belowground biomass of eelgrass in BB-LEH taken together for 2011 is highly problematic and reflective of an impacted coastal lagoon, even when considering only eelgrass in the central and south segments." Citations? As elsewhere in this section, I would have more confidence in the validation section if it included a prediction of eutrophic symptoms based on pressures and conditions, and a "goodness to fit" comparison of the actual symptoms with the predicted symptoms.
- Where data are available, some comparisons (summary tables with similar observations listed geographically?) of conditions reflecting the bay's decline to values from other coastal lagoonal systems would be helpful (e.g., Declines in eelgrass areal coverage, density, above and below ground biomass, and blade length have been documented to occur with increasing nutrient concentrations in other Mid-Attlantic ecosystems). We know that symptoms/conditions differ

among and within systems, but such comparisons would give readers more confidence in the stated conclusions.

Page 90:

- Synthesis and Management Recommendations - This is more a literature summary of impairments and eco-shifts already covered in the introduction due to human impacts and the need to develop a holistic plan to reduce these stressors. This section should describe what actions are needed to address the data gaps and uncertainties.

Pages 90-101:

- Component 5, Synthesis and Management Recommendations: The scope of work for this project specified that "thresholds levels of biotic decline and numeric nutrient loading criteria will be developed for the estuary and discussion of how these threshold levels can be integrated into a management plan will be given". On page 40, the report states that "recommendations for developing a management plan based on our findings will be given and additional data and analysis needed to improve the plan will be listed. The discussion that currently exists in this section needs to be expanded to adequately fulfill this task.

Page 91:

- Figure 5-2 should be revised to reflect the expression of nitrogen based on the selected threshold.
- Human factors should also include recreational usage, although not evaluated thru this project. Example –watercraft may be resuspending sediment and damaging SAV beds.

Page 92:

- This section should include some information generated by the USGS study completed as part of this project.
- "Based on a GIS analysis of the tidal marshes conducted by the Richard Stockton College Coastal Research Center, most of this wetland loss has occurred along the bay and tidal waterway shorelines." This was taken from the BBP 2011 SOTB Report, which should be correctly identified as the source. (This analysis was done at BBP request with BBP funding.)

Page 95:

- The report suggests that extensive macroalgal blooms were recorded and have persisted through ensuing years (2008-2010). However, Table 2-1 shows June-July 2008 had the highest percent macroalgae cover at 20% while for most other periods the coverage was less than 10%. Please provide a description and basis for the threshold used to conclude that macroalgal coverage is "extensive and persistent".

Page 97:

- "When TN loading exceeds some critical threshold value there is a triggering of phytoplankton and macroalgal blooms, as well as increased epiphytic growth, that can significantly reduce light transmission to seagrass beds, leading to acute die-offs of the seagrass and the resident shellfish and other benthic invertebrates inhabiting the beds." Please provide the threshold value nitrogen loading as well as the scientific basis for the selected threshold.

Page 98:

- The first paragraph indicates that the dissolved oxygen listing was based on continuous monitoring while the second paragraph indicates that this level of monitoring has not been done in Barnegat Bay. This conflict should be corrected.

- "BB-LEH Estuary is an impaired system both in respect to aquatic life support and human use as is evident in the conclusions of this study." This statement seems to contradict the 2011 validation period findings on page 83.
- "This listing for the north segment is based on continuous water quality monitoring…." Is this true? Please provide the continuous monitoring data that supports this claim. This section may need to be updated based upon current deliberations by NJDEP on the impaired status.

Page 99:

- Sea nettles are not used as an indicator of "impairment to swimmable waters" and are also not in NJ's assessment methodology for assessing primary contact recreation. Therefore, the following sentence in the 1st paragraph should be removed "High abundances of sea nettles have made bathing beaches and other waters in the estuary non-swimmable, creating impairment for human use". Similarly, 3rd paragraph should be deleted.
- This statement is not correct and should be removed. "The occurrence of sea nettle blooms in the north segment has resulted in extensive non-swimmable waters in violation of the Clean Water Act (Figure 5-6)." Sharks, alligators and sea nettles are not pollutants subject to restrictions under the Clean Water Act.
- "Sampling in 2011 had revealed much higher numbers of sea nettles at Brick (western side of Barnegat Bay) than Lavallette (eastern side of the Barnegat Bay) in the northern segment (Figure 5-5)."Again, please identify the BBP report as the source of the information.
- "The occurrence of sea nettle blooms in the north segment has resulted in extensive non-swimmable waters in violation of the Clean Water Act (Figure 5-6)."There is no standard or criterion for jellyfish, so this statement is incorrect. Not everyone may agree, but one potential modification would be to suggest that a standard/criterion for jellyfish be explored by state and federal regulatory agencies.

Page 100:

- 2nd to last paragraph on sea nettles, the last two sentences should be deleted: "….greater enforcement of regulations is necessary….long term solution is more effective administrative/management intervention".
- "Most importantly, much of the energy flow in food chains dominated by sea nettles does not pass upward to upper-trophic-level organisms, thereby reducing biotic production of the system." I believe Condon et al 2011 (?, or other Condon publications) would be appropriate citations to reference.
- Shellfish: "These numbers are indicative of an ongoing insidious ecological decline of the estuary. The cause of this dramatic decline has not been unequivocally established, although the diminution in hard clam landings has occurred during an escalating period of nutrient enrichment and eutrophication of the estuary. Hard clam landings are affected by several factors besides absolute abundance. For example, fishing effort, market value, and shellfish bed closures all affect hard clam harvest. Currently, BB-LEH has a very limited commercial fishery for hard clams, and it also has a limited recreational fishery." Without knowing the cause of the decline, I think it is inappropriate to make the first statement. I think it more appropriate to emphasize the need for additional shellfish research into the causes of shellfish decline (e.g., overfishing due to inadequate fishery management) and the resources needed to promote the shellfish industry.

Page 125:

- Figures 2-3 and 2-4 are not clear. Scale could be adjusted to make TP and TN concentrations reported over time more visible.

Page 140:

- Figure 3-2 shows that REMAP sediment TOC was used in the index but the section describing the indicators (pg. 60) doesn't include this indicator.

Page 182:

- Figures 5-5 and 5-6 should include the source of the sea nettle data.

Page 194:

- Table 3-2 doesn't include any information on the benthic invertebrate indicator and shows a reference of Baden et al. 1990 that isn't in the References chapter.

Page 202:

- Table 3-10 (Weightings of indicators) does not include the benthic invertebrate indicator.

Page 211:

- Baker et al. manuscript. This appears generally well done, but the entire manuscript needs to be better integrated into the overall report. Both the landscape and loadings information have considerable potential use in guiding next steps, in terms of research and implementation projects to address loadings. However, there do appear some notable priorities for research, which may not be answered by ongoing WQ monitoring in the watershed.

Page 221:

- Purpose and Scope (USGS portion): Should include the project objectives that were identified in the QAPP.

Page 237:

- With insufficient data in the southern segment, does this mean that we do not know if there is a good agreement between the measured and calculated values? Would this mean that there is a great level of uncertainty regarding the accuracy of the nitrogen loading estimates for the southern segment, since we have no way to compare to the observed data?

Page 238:

- Is there any concern on the accuracy of PLOAD due to the large over estimation by PLOAD during the calibration period? Could this also be due to an overestimation of the Land Use EMC's?

Page 239:

- Evaluation of available water quality data (USGS portion): Describe how the secondary data used for this project were evaluated. In particular, what temporal and spatial coverage criteria were used, if and how outliers were determined, how non-detects and missing values were used, and what the criteria were for age of data, comparable sampling and laboratory method criteria, etc. Representativeness of the secondary data also should have been evaluated and needs to be stated. This information should be summarized in a table along with source citations for the data.
- "Note: there were no sites with sufficient data in the southern segment." This statement would seem to severely compromise the eutrophication index and "bigger picture" conclusions of the overall report.

Page 240

- "…no significant difference between concentrations … during rising limb, peak, or falling limb…"Is this typical of sandy coastal watersheds? What are the implications for stormwater management?
- Please provide a reference, table or figure for this statement, "However, when averaged over many events over several years, there was no significant difference between concentrations of total nitrogen or orthophosphate during the rising limb, peak flow or falling limb of the discharge hydrographs for these three streams."
- Please explain how many days is several as stated in the sentence "...baseflow (less than 0.2 inches of precipitation for several days prior to sampling)."

Page 248:

- Summary and conclusions (USGS portion): Were there any management or technical recommendations from this portion of the study?

Page 295

- Sensitivity analysis: "Therefore, under the multi-year scenario (1999-2010) that includes total phosphorus, the weightings are: temperature 15%, dissolved oxygen 8%, total nitrogen 13%, and total phosphorus 65%." Does this indicate that the order of importance is total P, temperature, TN and DO? Therefore, TP is more important than TN in terms of the water quality component index? The importance of TP should be discussed in the text and weighed versus other conclusions on TN. A critical examination of the appropriateness of the sensitivity analysis in light of the multiple comments above (e.g., availability of data, weightings, etc) also needs to be completed. This examination should answer questions about whether the sensitivity analysis is robust or weak due to the variability in data availability.

Figure 3-39:

- Overall Eutrophication Index: A trend line should be fitted to each segment in order to evaluate the estimated direction/trend of the index. However, the feasibility and appropriateness of fitting a line should be considered after other comments have been addressed.

Appendix #1 (starting on page 211) and Appendix #3 (starting on page 324) were submitted with the report. Is there an Appendix #2?

Citations/references:

- Page 27 - text includes citation: " Velinsky et. al., 2010", while the reference section (on page 113) lists " Velinsky et. al., 2011". Please clarify.
- Page 40 - citation EPA-822-B-01-003, 2001 is not listed in the references section
- Pages 35, 41 and 233 - citation "USEPA, 2001" is not listed in references section
- Page 65 - citation "Deegan et al., 2002" are not listed in references section
- Pages 65 and 147 - citation "Tomasko et al., 1996" is not listed in references section
- Page 66 - citation "Burkholder et al., 2001" is not listed in references section
- Page 67 - citations "Weisberg et al., 1997", " Van Dolah et al., 1999", and " Hale and Heltshe, 2008" are not listed in references section